Data Collection Summary:

The dataset for this project is a multi-sheet Excel file downloaded from Kaggle. The dataset is related to the domain of Supply Chain Management, which aligns with my professional background.

The dataset comprises seven tables in total. One of these tables, 'OrderList', contains historical data on how orders were assigned routes in the past. This table will serve as the primary source of data for the project.

The remaining six tables ('FreightRates', 'WhCosts', 'WhCapacities', 'ProductsPerPlant', 'VmiCustomers', 'PlantPorts') provide additional information about various restrictions and constraints that must be considered when assigning routes to orders. These restrictions include specific customer-plant associations, physical connectivity between plants and ports, and specific item handling capabilities of plants.

Problem Statement:

The goal of this Machine Learning and Data Science capstone project is to develop a model that can effectively assign routes to orders listed in the 'OrderList' table, given the various restrictions outlined in the other six tables. The model should consider constraints such as weight restrictions, specific customer-plant servicing rules, physical plant-port connections, and specific item handling capabilities of plants.

The ultimate objective is to minimize the overall costs associated with freight and warehousing while ensuring all orders are shipped to their respective destinations. This is a deterministic optimization problem focusing on orders from a specific day.

The success of this project will be evaluated based on the model's ability to minimize costs while adhering to all given constraints. The historical data provided in the 'OrderList' table will serve as a valuable resource for understanding past strategies and informing the development of the predictive model.